

# OpenMP Tutorial



## A “Hands-on” Introduction to OpenMP\*

Larry Meadows  
Intel Corporation  
lawrence.f.meadows@intel.com

Mark Bull  
EPCC  
markb@epcc.ed.ac.uk

Tim Mattson  
Intel Corporation  
timothy.g.mattson@intel.com

\* The name “OpenMP” is the property of the OpenMP Architecture Review Board.

1

## Preliminaries: part 1

- Disclosures
  - ◆ The views expressed in this tutorial are those of the people delivering the tutorial.
    - We are not speaking for our employers.
    - We are not speaking for the OpenMP ARB
- We take these tutorials VERY seriously:
  - ◆ Help us improve ... tell us how you would make this tutorial better.

2

# OpenMP Tutorial

## Preliminaries: Part 2

- Our plan for the day .. Active learning!
  - ◆ We will mix short lectures with short exercises.
  - ◆ You will use your laptop for the exercises ... that way you'll have an OpenMP environment to take home so you can keep learning on your own.
- Please follow these simple rules
  - ◆ Do the exercises we assign and then change things around and experiment.
    - Embrace active learning!
  - ◆ Don't cheat: Do Not look at the solutions before you complete an exercise ... even if you get really frustrated.

3

## Our Plan for the day

Topic	Exercise	concepts
I. OMP Intro	Install sw, hello_world	Parallel regions
II. Creating threads	Pi_spmc_simple	Parallel, default data environment, runtime library calls
III. Synchronization	Pi_spmc_final	False sharing, critical, atomic
IV. Parallel loops	Pi_loop, Matmul	For, schedule, reduction,
V. Odds and ends	No exercise	Single, sections, master, runtime libraries, environment variables, synchronization, etc.
VI. Data Environment	Molecular Dyn.	Data environment details, software optimization
VII. OpenMP 3 and tasks	Linked list (tasks) Linked list (no tasks)	Tasks and other OpenMP 3 features
VIII. Memory model	Producer consumer	Point to point synch with flush
IX. Threadprivate	Pi_mc	Modular software

4

# OpenMP Tutorial

## Outline

- ➔ ● Introduction to OpenMP
- Creating Threads
- Synchronization
- Parallel Loops
- Synchronize single masters and stuff
- Data environment
- OpenMP 3.0 and Tasks
- Memory model
- Threadprivate Data

5

## OpenMP\* Overview:

### *OpenMP: An API for Writing Multithreaded Applications*

- A set of compiler directives and library routines for parallel application programmers
- Greatly simplifies writing multi-threaded (MT) programs in Fortran, C and C++
- Standardizes last 20 years of SMP practice

```
C$OMP PARALLEL COPYIN(/blk/)
```

```
C$OMP DO lastprivate(XX)
```

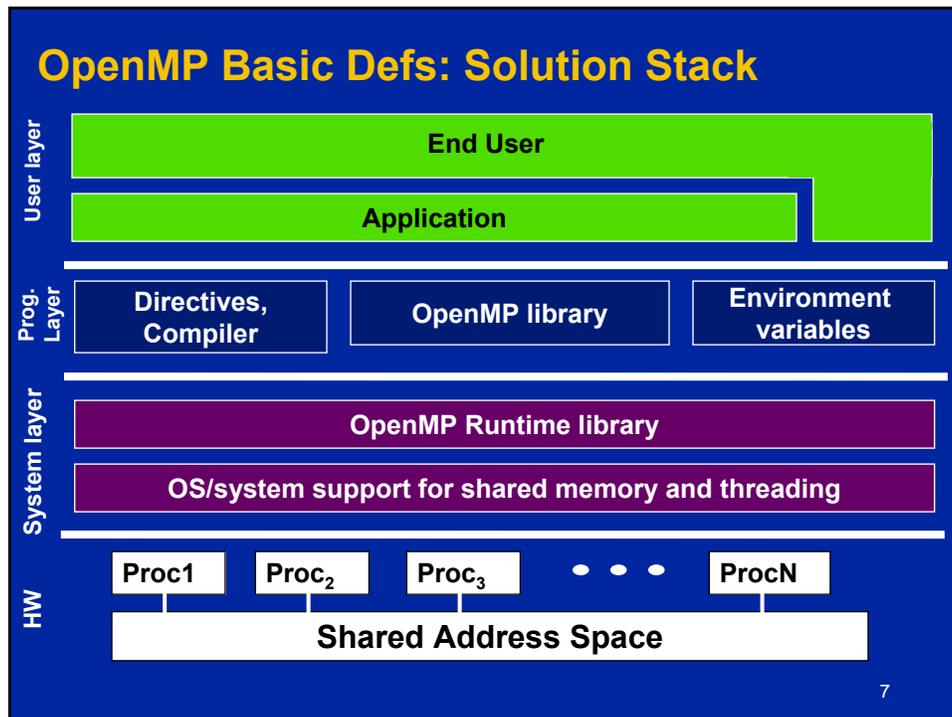
```
Nthrds = OMP_GET_NUM_PROCS()
```

```
omp_set_lock(lck)
```

\* The name "OpenMP" is the property of the OpenMP Architecture Review Board.

6

# OpenMP Tutorial



## OpenMP core syntax

- Most of the constructs in OpenMP are compiler directives.
  - ◆ `#pragma omp construct [clause [clause]...]`
  - ◆ Example
    - ◆ `#pragma omp parallel num_threads(4)`
- Function prototypes and types in the file:
  - ◆ `#include <omp.h>`
- Most OpenMP\* constructs apply to a “structured block”.
  - ◆ Structured block: a block of one or more statements with one point of entry at the top and one point of exit at the bottom.
  - ◆ It’s OK to have an `exit()` within the structured block.

8

# OpenMP Tutorial

## Exercise 1, Part A: Hello world

Verify that your environment works

- Write a program that prints “hello world”.

```
void main()
{

    int ID = 0;
    printf(" hello(%d) ", ID);
    printf(" world(%d) \n", ID);

}
```

9

## Exercise 1, Part B: Hello world

Verify that your OpenMP environment works

- Write a multithreaded program that prints “hello world”.

```
#include "omp.h"
void main()
{
    #pragma omp parallel
    {
        int ID = 0;
        printf(" hello(%d) ", ID);
        printf(" world(%d) \n", ID);
    }
}
```

Switches for compiling and linking

-fopenmp gcc

-mp pgi

/Qopenmp intel

10

# OpenMP Tutorial

## Exercise 1: Solution

### A multi-threaded “Hello world” program

- Write a multithreaded program where each thread prints “hello world”.

```
#include "omp.h"
void main()
{
#pragma omp parallel
{
    int ID = omp_get_thread_num();
    printf(" hello(%d) ", ID);
    printf(" world(%d) \n", ID);
}
}
```

Annotations:

- OpenMP include file (points to `#include "omp.h"`)
- Parallel region with default number of threads (points to `#pragma omp parallel`)
- Runtime library function to return a thread ID. (points to `omp_get_thread_num()`)
- End of the Parallel region (points to the closing brace of the parallel region)

**Sample Output:**

```
hello(1) hello(0) world(1)
world(0)
hello (3) hello(2) world(3)
world(2)
```

11

## OpenMP Overview: How do threads interact?

- OpenMP is a multi-threading, shared address model.
  - Threads communicate by sharing variables.
- Unintended sharing of data causes race conditions:
  - race condition: when the program's outcome changes as the threads are scheduled differently.
- To control race conditions:
  - Use synchronization to protect data conflicts.
- Synchronization is expensive so:
  - Change how data is accessed to minimize the need for synchronization.

12

# OpenMP Tutorial

## Outline

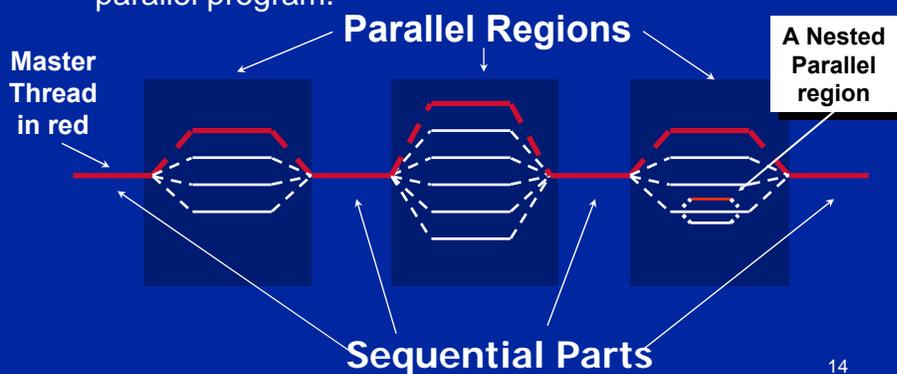
- Introduction to OpenMP
- ➔ • Creating Threads
- Synchronization
- Parallel Loops
- Synchronize single masters and stuff
- Data environment
- OpenMP 3.0 and Tasks
- Memory model
- Threadprivate Data

13

## OpenMP Programming Model:

### Fork-Join Parallelism:

- ◆ Master thread spawns a team of threads as needed.
- ◆ Parallelism added incrementally until performance goals are met: i.e. the sequential program evolves into a parallel program.



14

# OpenMP Tutorial

## Thread Creation: Parallel Regions

- You create threads in OpenMP\* with the parallel construct.
- For example, To create a 4 thread Parallel region:

Each thread executes a copy of the code within the structured block

```
double A[1000];  
omp_set_num_threads(4);  
#pragma omp parallel  
{  
    int ID = omp_get_thread_num();  
    pooh(ID,A);  
}
```

Runtime function to request a certain number of threads

Runtime function returning a thread ID

- Each thread calls `pooh(ID,A)` for `ID = 0 to 3`

15

\* The name "OpenMP" is the property of the OpenMP Architecture Review Board

## Thread Creation: Parallel Regions

- You create threads in OpenMP\* with the parallel construct.
- For example, To create a 4 thread Parallel region:

Each thread executes a copy of the code within the structured block

```
double A[1000];  
#pragma omp parallel num_threads(4)  
{  
    int ID = omp_get_thread_num();  
    pooh(ID,A);  
}
```

clause to request a certain number of threads

Runtime function returning a thread ID

- Each thread calls `pooh(ID,A)` for `ID = 0 to 3`

16

\* The name "OpenMP" is the property of the OpenMP Architecture Review Board

# OpenMP Tutorial

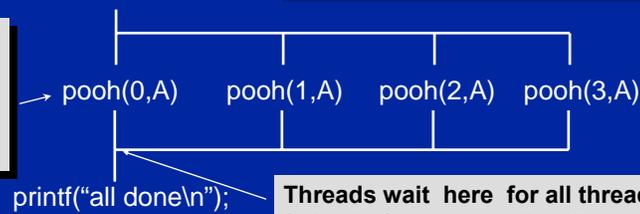
## Thread Creation: Parallel Regions example

- Each thread executes the same code redundantly.

```
double A[1000];  
omp_set_num_threads(4)
```

```
double A[1000];  
omp_set_num_threads(4);  
#pragma omp parallel  
{  
    int ID = omp_get_thread_num();  
    pooh(ID, A);  
}  
printf("all done\n");
```

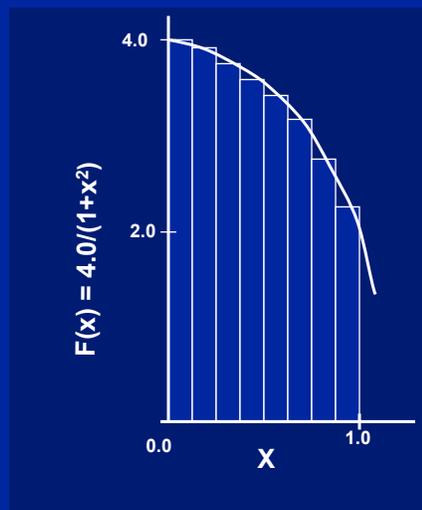
A single copy of A is shared between all threads.



Threads wait here for all threads to finish before proceeding (i.e. a *barrier*)

\* The name "OpenMP" is the property of the OpenMP Architecture Review Board

## Exercises 2 to 4: Numerical Integration



Mathematically, we know that:

$$\int_0^1 \frac{4.0}{(1+x^2)} dx = \pi$$

We can approximate the integral as a sum of rectangles:

$$\sum_{i=0}^N F(x_i) \Delta x \approx \pi$$

Where each rectangle has width  $\Delta x$  and height  $F(x_i)$  at the middle of interval  $i$ .

18

# OpenMP Tutorial

## Exercises 2 to 4: Serial PI Program

```
static long num_steps = 100000;
double step;
void main ()
{   int i;  double x, pi, sum = 0.0;

    step = 1.0/(double) num_steps;

    for (i=0;i< num_steps; i++){
        x = (i+0.5)*step;
        sum = sum + 4.0/(1.0+x*x);
    }
    pi = step * sum;
}
```

19

## Exercise 2

- Create a parallel version of the pi program using a parallel construct.
- Pay close attention to shared versus private variables.
- In addition to a parallel construct, you will need the runtime library routines

◆ `int omp_get_num_threads();`

Number of threads in the team

◆ `int omp_get_thread_num();`

Thread ID or rank

◆ `double omp_get_wtime();`

Time in Seconds since a fixed point in the past

20



# OpenMP Tutorial

## Outline

- Introduction to OpenMP
- Creating Threads
- ➔ • Synchronization
- Parallel Loops
- Synchronize single masters and stuff
- Data environment
- OpenMP 3.0 and Tasks
- Memory model
- Threadprivate Data

21

## Synchronization

- High level synchronization:

- critical
- atomic

- barrier
- ordered

- Low level synchronization

- flush
- locks (both simple and nested)

Synchronization is used to impose order constraints and to protect access to shared data

Discussed later

22

# OpenMP Tutorial

## Synchronization: critical

- **Mutual exclusion:** Only one thread at a time can enter a **critical** region.

Threads wait their turn – only one at a time calls consume()

```
float res;
#pragma omp parallel
{
    float B; int i, id, nthrds;
    id = omp_get_thread_num();
    nthrds = omp_get_num_threads();
    for(i=id;i<niters;i+nthrds){
        B = big_job(i);
        #pragma omp critical
            consume (B, res);
    }
}
```

## Synchronization: Atomic

- **Atomic** provides mutual exclusion but only applies to the update of a memory location (the update of X in the following example)

```
#pragma omp parallel
{
    double tmp, B;
    B = DOIT();
    tmp = big_ugly(B);
    #pragma omp atomic
        X += tmp;
}
```

Atomic only protects the read/update of X

24

# OpenMP Tutorial

## Exercise 3

- In exercise 2, you probably used an array to create space for each thread to store its partial sum.
- If array elements happen to share a cache line, this leads to false sharing.
  - Non-shared data in the same cache line so each update invalidates the cache line ... in essence “sloshing independent data” back and forth between threads.
- Modify your “pi program” from exercise 2 to avoid false sharing due to the sum array.

25



## Outline

- Introduction to OpenMP
- Creating Threads
- Synchronization
- ➔ ● Parallel Loops
- Synchronize single masters and stuff
- Data environment
- OpenMP 3.0 and Tasks
- Memory model
- Threadprivate Data

26

# OpenMP Tutorial

## SPMD vs. worksharing

- A parallel construct by itself creates an SPMD or “Single Program Multiple Data” program ... i.e., each thread redundantly executes the same code.
- How do you split up pathways through the code between threads within a team?
  - ◆ This is called worksharing
    - Loop construct
    - Sections/section constructs
    - Single construct
    - Task construct .... Coming in OpenMP 3.0

Discussed later

27

## The loop worksharing Constructs

- The loop worksharing construct splits up loop iterations among the threads in a team

```
#pragma omp parallel
{
#pragma omp for
  for (l=0;l<N;l++){
    NEAT_STUFF(l);
  }
}
```

Loop construct name:

- C/C++: for
- Fortran: do

The variable l is made “private” to each thread by default. You could do this explicitly with a “private(l)” clause

28

# OpenMP Tutorial

## Loop worksharing Constructs A motivating example

Sequential code

```
for(i=0;i<N;i++) { a[i] = a[i] + b[i];}
```

OpenMP parallel region

```
#pragma omp parallel
{
    int id, i, Nthrds, istart, iend;
    id = omp_get_thread_num();
    Nthrds = omp_get_num_threads();
    istart = id * N / Nthrds;
    iend = (id+1) * N / Nthrds;
    if (id == Nthrds-1)iend = N;
    for(i=istart;i<iend;i++) { a[i] = a[i] + b[i];}
}
```

OpenMP parallel region and a worksharing for construct

```
#pragma omp parallel
#pragma omp for
for(i=0;i<N;i++) { a[i] = a[i] + b[i];}
```

## loop worksharing constructs: The schedule clause

- The schedule clause affects how loop iterations are mapped onto threads
  - ◆ `schedule(static [,chunk])`
    - Deal-out blocks of iterations of size “chunk” to each thread.
  - ◆ `schedule(dynamic[,chunk])`
    - Each thread grabs “chunk” iterations off a queue until all iterations have been handled.
  - ◆ `schedule(guided[,chunk])`
    - Threads dynamically grab blocks of iterations. The size of the block starts large and shrinks down to size “chunk” as the calculation proceeds.
  - ◆ `schedule(runtime)`
    - Schedule and chunk size taken from the OMP\_SCHEDULE environment variable (or the runtime library ... for OpenMP 3.0).

30

# OpenMP Tutorial

## loop work-sharing constructs: The schedule clause

Schedule Clause	When To Use
STATIC	Pre-determined and predictable by the programmer
DYNAMIC	Unpredictable, highly variable work per iteration
GUIDED	Special case of dynamic to reduce scheduling overhead

Least work at runtime : scheduling done at compile-time

Most work at runtime : complex scheduling logic used at run-time

31

## Combined parallel/worksharing construct

- OpenMP shortcut: Put the “parallel” and the worksharing directive on the same line

```
double res[MAX]; int i;
#pragma omp parallel
{
    #pragma omp for
    for (i=0; i< MAX; i++) {
        res[i] = huge();
    }
}
```

```
double res[MAX]; int i;
#pragma omp parallel for
for (i=0; i< MAX; i++) {
    res[i] = huge();
}
```

These are equivalent

32

# OpenMP Tutorial

## Working with loops

- Basic approach
  - ◆ Find compute intensive loops
  - ◆ Make the loop iterations independent .. So they can safely execute in any order without loop-carried dependencies
  - ◆ Place the appropriate OpenMP directive and test

```
int i, j, A[MAX];
j = 5;
for (i=0; i< MAX; i++) {
    j += 2;
    A[i] = big(j);
}
```

Note: loop index  
"i" is private by  
default

Remove loop  
carried  
dependence

```
int i, A[MAX];
#pragma omp parallel for
for (i=0; i< MAX; i++) {
    int j = 5 + 2*i;
    A[i] = big(j);
}
```

33

## Reduction

- How do we handle this case?

```
double ave=0.0, A[MAX]; int i;
for (i=0; i< MAX; i++) {
    ave += A[i];
}
ave = ave/MAX;
```

- We are combining values into a single accumulation variable (ave) ... there is a true dependence between loop iterations that can't be trivially removed
- This is a very common situation ... it is called a "reduction".
- Support for reduction operations is included in most parallel programming environments.

34

# OpenMP Tutorial

## Reduction

- OpenMP reduction clause:  
**reduction (op : list)**
- Inside a parallel or a work-sharing construct:
  - A local copy of each list variable is made and initialized depending on the “op” (e.g. 0 for “+”).
  - Updates occur on the local copy.
  - Local copies are reduced into a single value and combined with the original global value.
- The variables in “list” must be shared in the enclosing parallel region.

```
double ave=0.0, A[MAX]; int i;
#pragma omp parallel for reduction (+:ave)
for (i=0;i< MAX; i++) {
    ave + = A[i];
}
ave = ave/MAX;
```

35

## OpenMP: Reduction operands/initial-values

- Many different associative operands can be used with reduction:
- Initial values are the ones that make sense mathematically.

Operator	Initial value
+	0
*	1
-	0

C/C++ only	
Operator	Initial value
&	~0
	0
^	0
&&	1
	0

Fortran Only	
Operator	Initial value
.AND.	.true.
.OR.	.false.
.NEQV.	.false.
.IEOR.	0
.IOR.	0
.IAND.	All bits on
.EQV.	.true.
MIN*	Largest pos. number
MAX*	Most neg. number

36

# OpenMP Tutorial

## Exercise 4: Pi with loops

- Go back to the serial pi program and parallelize it with a loop construct
- Your goal is to minimize the number of changes made to the serial program.

37



## Exercise 5: Optimizing loops

- Parallelize the matrix multiplication program in the file matmul.c
- Can you optimize the program by playing with how the loops are scheduled?

38



# OpenMP Tutorial

## Outline

- Introduction to OpenMP
- Creating Threads
- Synchronization
- Parallel Loops
- ➔ • Synchronize single masters and stuff
- Data environment
- OpenMP 3.0 and Tasks
- Memory model
- Threadprivate Data

39

## Synchronization: Barrier

- **Barrier**: Each thread waits until all threads arrive.

```
#pragma omp parallel shared (A, B, C) private(id)
{
    id=omp_get_thread_num();
    A[id] = big_calc1(id);
    #pragma omp barrier
    #pragma omp for
        for(i=0;i<N;i++){C[i]=big_calc3(i,A);}
    #pragma omp for nowait
        for(i=0;i<N;i++){ B[i]=big_calc2(C, i); }
    A[id] = big_calc4(id);
}
```

implicit barrier at the end of a for worksharing construct

implicit barrier at the end of a parallel region

no implicit barrier due to nowait

# OpenMP Tutorial

## Master Construct

- The **master** construct denotes a structured block that is only executed by the master thread.
- The other threads just skip it (no synchronization is implied).

```
#pragma omp parallel
{
    do_many_things();
    #pragma omp master
    { exchange_boundaries(); }
    #pragma omp barrier
    do_many_other_things();
}
```

41

## Sections worksharing Construct

- The **Sections worksharing** construct gives a different structured block to each thread.

```
#pragma omp parallel
{
    #pragma omp sections
    {
        #pragma omp section
        X_calculation();
        #pragma omp section
        y_calculation();
        #pragma omp section
        z_calculation();
    }
}
```

By default, there is a barrier at the end of the “omp sections”. Use the “nowait” clause to turn off the barrier.

42

# OpenMP Tutorial

## Single worksharing Construct

- The **single** construct denotes a block of code that is executed by only one thread (not necessarily the master thread).
- A barrier is implied at the end of the single block (can remove the barrier with a *nowait* clause).

```
#pragma omp parallel
{
    do_many_things();
    #pragma omp single
    { exchange_boundaries(); }
    do_many_other_things();
}
```

43

## Synchronization: ordered

- The **ordered** region executes in the sequential order.

```
#pragma omp parallel private (tmp)
#pragma omp for ordered reduction(+:res)
    for (l=0;l<N;l++){
        tmp = NEAT_STUFF(l);
        #pragma ordered
        res += consum(tmp);
    }
```

44

# OpenMP Tutorial

## Synchronization: Lock routines

- Simple Lock routines:

- ◆ A simple lock is available if it is unset.

- `omp_init_lock()`, `omp_set_lock()`,  
`omp_unset_lock()`, `omp_test_lock()`,  
`omp_destroy_lock()`

A lock implies a memory fence (a “flush”) of all thread visible variables

- Nested Locks

- ◆ A nested lock is available if it is unset or if it is set but owned by the thread executing the nested lock function

- `omp_init_nest_lock()`, `omp_set_nest_lock()`,  
`omp_unset_nest_lock()`, `omp_test_nest_lock()`,  
`omp_destroy_nest_lock()`

Note: a thread always accesses the most recent copy of the lock, so you don't need to use a flush on the lock variable.

45

## Synchronization: Simple Locks

- Protect resources with locks.

```
omp_lock_t lck;  
omp_init_lock(&lck);  
#pragma omp parallel private (tmp, id)  
{  
    id = omp_get_thread_num();  
    tmp = do_lots_of_work(id);  
    omp_set_lock(&lck);  
    printf(“%d %d”, id, tmp);  
    omp_unset_lock(&lck);  
}  
omp_destroy_lock(&lck);
```

Wait here for your turn.

Release the lock so the next thread gets a turn.

Free-up storage when done.

# OpenMP Tutorial

## Runtime Library routines

- Runtime environment routines:
  - Modify/Check the number of threads
    - `omp_set_num_threads()`, `omp_get_num_threads()`,  
`omp_get_thread_num()`, `omp_get_max_threads()`
  - Are we in an active parallel region?
    - `omp_in_parallel()`
  - Do you want the system to dynamically vary the number of threads from one parallel construct to another?
    - `omp_set_dynamic`, `omp_get_dynamic()`;
  - How many processors in the system?
    - `omp_num_procs()`

...plus a few less commonly used routines.

47

## Runtime Library routines

- To use a known, fixed number of threads in a program, (1) tell the system that you don't want dynamic adjustment of the number of threads, (2) set the number of threads, then (3) save the number you got.

```
#include <omp.h>
void main()
{ int num_threads;
  omp_set_dynamic( 0 );
  omp_set_num_threads( omp_num_procs() );
#pragma omp parallel
  { int id=omp_get_thread_num();
#pragma omp single
  num_threads = omp_get_num_threads();
  do_lots_of_stuff(id);
  }
}
```

Disable dynamic adjustment of the number of threads.

Request as many threads as you have processors.

Protect this op since Memory stores are not atomic

Even in this case, the system may give you fewer threads than requested. If the precise # of threads matters, test for it and respond accordingly.

48

# OpenMP Tutorial

## Environment Variables

- Set the default number of threads to use.
  - `OMP_NUM_THREADS` *int\_literal*
- Control how “omp for schedule(RUNTIME)” loop iterations are scheduled.
  - `OMP_SCHEDULE` “schedule[, chunk\_size]”

... Plus several less commonly used environment variables.

49

## Outline

- Introduction to OpenMP
- Creating Threads
- Synchronization
- Parallel Loops
- Synchronize single masters and stuff
- ➡ ● Data environment
- OpenMP 3.0 and Tasks
- Memory model
- Threadprivate Data

50

# OpenMP Tutorial

## Data environment: Default storage attributes

- Shared Memory programming model:
  - Most variables are shared by default
- Global variables are SHARED among threads
  - Fortran: COMMON blocks, SAVE variables, MODULE variables
  - C: File scope variables, static
  - Both: dynamically allocated memory (ALLOCATE, malloc, new)
- But not everything is shared...
  - Stack variables in subprograms(Fortran) or functions(C) called from parallel regions are PRIVATE
  - Automatic variables within a statement block are PRIVATE.

51

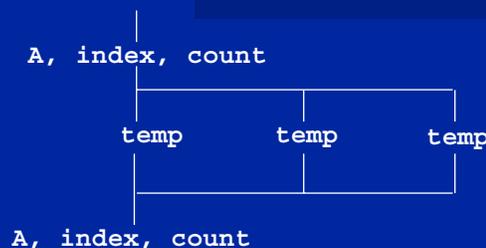
## Data sharing: Examples

```
double A[10];
int main() {
  int index[10];
  #pragma omp parallel
    work(index);
  printf("%d\n", index[0]);
}
```

A, index and count are shared by all threads.

temp is local to each thread

```
extern double A[10];
void work(int *index) {
  double temp[10];
  static int count;
  ...
}
```



52

\* Third party trademarks and names are the property of their respective owner.

# OpenMP Tutorial

## Data sharing: Changing storage attributes

- One can selectively change storage attributes for constructs using the following clauses\*
  - SHARED
  - PRIVATE
  - FIRSTPRIVATE
- The final value of a private inside a parallel loop can be transmitted to the shared variable outside the loop with:
  - LASTPRIVATE
- The default attributes can be overridden with:
  - DEFAULT (PRIVATE | SHARED | NONE)  
DEFAULT(PRIVATE) *is Fortran only*

All the clauses on this page apply to the OpenMP construct NOT to the entire region.

All data clauses apply to parallel constructs and worksharing constructs except "shared" which only applies to parallel constructs.

53

## Data Sharing: Private Clause

- `private(var)` creates a new local copy of `var` for each thread.
  - The value is uninitialized
  - In OpenMP 2.5 the value of the shared variable is undefined after the region

```
void wrong() {  
    int tmp = 0;  
    #pragma omp for private(tmp)  
    for (int j = 0; j < 1000; ++j)  
        tmp += j;  
    printf("%d\n", tmp);  
}
```

tmp was not initialized

tmp: 0 in 3.0,  
unspecified in 2.5

54

# OpenMP Tutorial

## Data Sharing: Private Clause When is the original variable valid?

- The original variable's value is unspecified in OpenMP 2.5.
- In OpenMP 3.0, if it is referenced outside of the construct
  - Implementations may reference the original variable or a copy .....
  - A dangerous programming practice!

```
int tmp;  
void danger() {  
    tmp = 0;  
    #pragma omp parallel private(tmp)  
    work();  
    printf("%d\n", tmp);  
}
```

tmp has unspecified value

```
extern int tmp;  
void work() {  
    tmp = 5;  
}
```

unspecified which copy of tmp

55

## Data Sharing: Firstprivate Clause

- Firstprivate is a special case of private.
  - Initializes each private copy with the corresponding value from the master thread.

```
void useless() {  
    int tmp = 0;  
    #pragma omp for firstprivate(tmp)  
    for (int j = 0; j < 1000; ++j)  
        tmp += j;  
    printf("%d\n", tmp);  
}
```

Each thread gets its own tmp with an initial value of 0

tmp: 0 in 3.0, unspecified in 2.5

56

# OpenMP Tutorial

## Data sharing: Lastprivate Clause

- Lastprivate passes the value of a private from the last iteration to a global variable.

```
void closer() {  
    int tmp = 0;  
    #pragma omp parallel for firstprivate(tmp) \  
    lastprivate(tmp)  
    for (int j = 0; j < 1000; ++j)  
        tmp += j;  
    printf("%d\n", tmp);  
}
```

Each thread gets its own tmp with an initial value of 0

tmp is defined as its value at the "last sequential" iteration (i.e., for j=999)

57

## Data Sharing: A data environment test

- Consider this example of PRIVATE and FIRSTPRIVATE

```
variables A,B, and C = 1  
#pragma omp parallel private(B) firstprivate(C)
```

- Are A,B,C local to each thread or shared inside the parallel region?
- What are their initial values inside and values after the parallel region?

### Inside this parallel region ...

- "A" is shared by all threads; equals 1
- "B" and "C" are local to each thread.
  - B's initial value is undefined
  - C's initial value equals 1

### Outside this parallel region ...

- The values of "B" and "C" are unspecified in OpenMP 2.5, and in OpenMP 3.0 if referenced in the region but outside the construct.

58

# OpenMP Tutorial

## Data Sharing: Default Clause

- Note that the default storage attribute is **DEFAULT(SHARED)** (so no need to use it)
  - ◆ Exception: **#pragma omp task**
- To change default: **DEFAULT(PRIVATE)**
  - ◆ *each* variable in the construct is made private as if specified in a private clause
  - ◆ mostly saves typing
- **DEFAULT(NONE)**: *no* default for variables in static extent. Must list storage attribute for each variable in static extent. Good programming practice!

Only the Fortran API supports default(private).

C/C++ only has default(shared) or default(none).

59

## Data Sharing: Default Clause Example

```
itotal = 1000
C$OMP PARALLEL PRIVATE(np, each)
  np = omp_get_num_threads()
  each = itotal/np
  .....
C$OMP END PARALLEL
```

These two  
code  
fragments are  
equivalent

```
itotal = 1000
C$OMP PARALLEL DEFAULT(PRIVATE) SHARED(itotal)
  np = omp_get_num_threads()
  each = itotal/np
  .....
C$OMP END PARALLEL
```

60

# OpenMP Tutorial

OpenMP 3.0

## Data Sharing: tasks (OpenMP 3.0)

- The default for tasks is usually `firstprivate`, because the task may not be executed until later (and variables may have gone out of scope).
- Variables that are shared in all constructs starting from the innermost enclosing parallel construct are shared, because the barrier guarantees task completion.

```
#pragma omp parallel shared(A) private(B)
{
  ...
  #pragma omp task
  {
    int C;
    compute(A, B, C);
  }
}
```

A is shared  
B is firstprivate  
C is private

61

## Exercise 6: Molecular dynamics

- The code supplied is a simple molecular dynamics simulation of the melting of solid argon.
- Computation is dominated by the calculation of force pairs in subroutine `forces` (in `forces.c`)
- Parallelise this routine using a parallel for construct and atomics. Think carefully about which variables should be `SHARED`, `PRIVATE` or `REDUCTION` variables.
- Experiment with different schedules kinds.

62

# OpenMP Tutorial

## Exercise 6 (cont.)

- Once you have a working version, move the parallel region out to encompass the iteration loop in main.c
  - ◆ code other than the forces loop must be executed by a single thread (or workshared).
  - ◆ how does the data sharing change?
- The atomics are a bottleneck on most systems.
  - ◆ This can be avoided by introducing a temporary array for the force accumulation, with an extra dimension indexed by thread number.
  - ◆ Which thread(s) should do the final accumulation into f?

63



## Outline

- Introduction to OpenMP
- Creating Threads
- Synchronization
- Parallel Loops
- Synchronize single masters and stuff
- Data environment
- ➔ ● OpenMP 3.0 and Tasks
- Memory model
- Threadprivate Data

64

# OpenMP Tutorial

## OpenMP pre-history

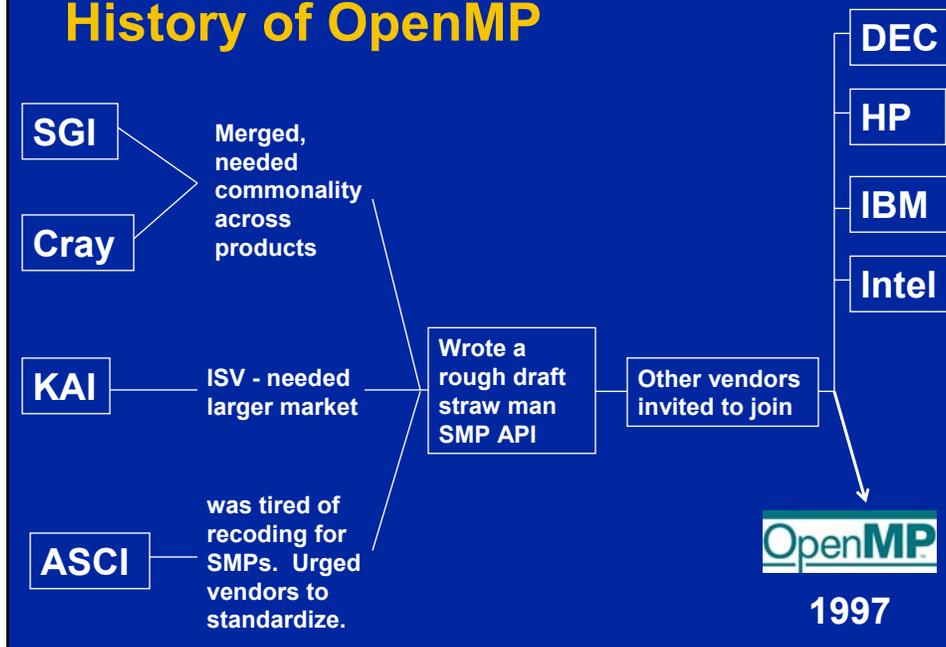
- OpenMP based upon SMP directive standardization efforts PCF and aborted ANSI X3H5 – late 80's
  - ◆ Nobody fully implemented either standard
  - ◆ Only a couple of partial implementations
- Vendors considered proprietary API's to be a competitive feature:
  - ◆ Every vendor had proprietary directives sets
  - ◆ Even KAP, a “portable” multi-platform parallelization tool used different directives on each platform

PCF – Parallel computing forum

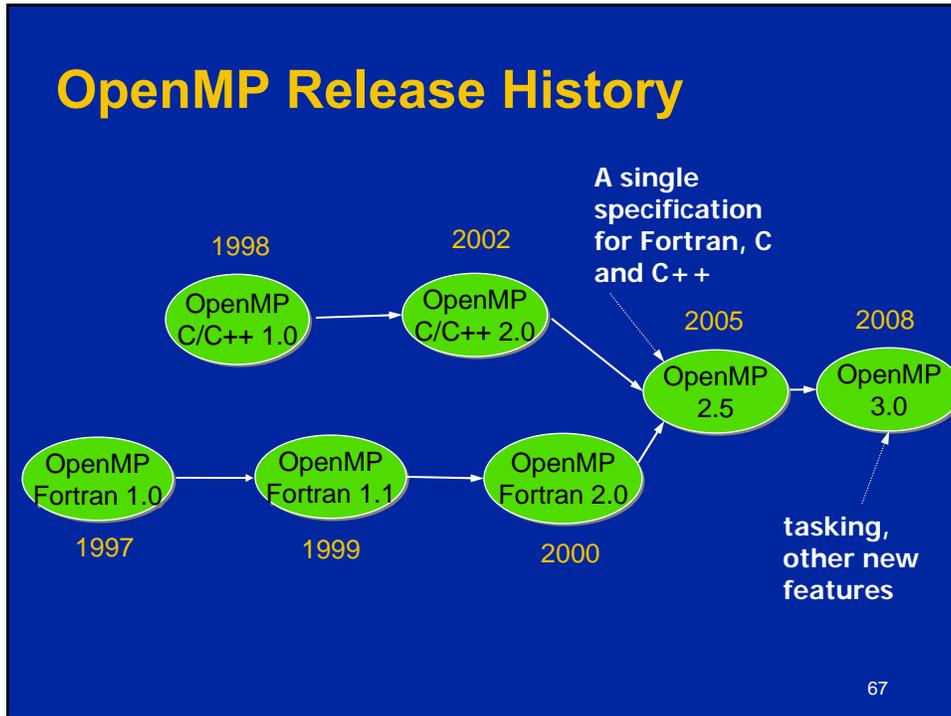
KAP – parallelization tool from KAI.

65

## History of OpenMP



# OpenMP Tutorial



## Tasks

OpenMP 3.0

- Adding tasking is the biggest addition for 3.0
- Worked on by a separate subcommittee
  - ◆ led by Jay Hoeflinger at Intel
- Re-examined issue from ground up
  - ◆ quite different from Intel taskq's

68

# OpenMP Tutorial

OpenMP 3.0

## General task characteristics

- A task has
  - ◆ Code to execute
  - ◆ A data environment (it *owns* its data)
  - ◆ An assigned thread that executes the code and uses the data
- Two activities: packaging and execution
  - ◆ Each encountering thread packages a new instance of a task (code and data)
  - ◆ Some thread in the team executes the task at some later time

69

OpenMP 3.0

## Definitions

- **Task construct** – `task` directive plus structured block
- **Task** – the package of code and instructions for allocating data created when a thread encounters a task construct
- **Task region** – the dynamic sequence of instructions produced by the execution of a task by a thread

70

# OpenMP Tutorial

OpenMP 3.0

## Tasks and OpenMP

- Tasks have been fully integrated into OpenMP
- Key concept: OpenMP has always had tasks, we just never called them that.
  - ◆ Thread encountering `parallel` construct packages up a set of *implicit* tasks, one per thread.
  - ◆ Team of threads is created.
  - ◆ Each thread in team is assigned to one of the tasks (and *tied* to it).
  - ◆ Barrier holds original master thread until all implicit tasks are finished.
- We have simply added a way to create a task explicitly for the team to execute.
- Every part of an OpenMP program is part of one task or another!

71

OpenMP 3.0

## task Construct

```
#pragma omp task [clause[[,clause] ...]  
    structured-block
```

where *clause* can be one of:

```
    if (expression)  
    untied  
    shared (list)  
    private (list)  
    firstprivate (list)  
    default( shared | none )
```

72

# OpenMP Tutorial

OpenMP 3.0

## The `if` clause

- When the `if` clause argument is false
  - ◆ The task is executed immediately by the encountering thread.
  - ◆ The data environment is still local to the new task...
  - ◆ ...and it's still a different task with respect to synchronization.
- It's a user directed optimization
  - ◆ when the cost of deferring the task is too great compared to the cost of executing the task code
  - ◆ to control cache and memory affinity

73

OpenMP 3.0

## When/where are tasks complete?

- At thread barriers, explicit or implicit
  - ◆ applies to all tasks generated in the current parallel region up to the barrier
  - ◆ matches user expectation
- At task barriers
  - ◆ i.e. Wait until all tasks defined in the current task have completed.  
`#pragma omp taskwait`
  - ◆ Note: applies only to tasks generated in the current task, not to "descendants" .

74

# OpenMP Tutorial

OpenMP 3.0

## Example – parallel pointer chasing using tasks

```
#pragma omp parallel
{
  #pragma omp single private(p)
  {
    p = listhead ;
    while (p) {
      #pragma omp task
      process (p);
      p=next (p) ;
    }
  }
}
```

p is firstprivate inside this task

75

OpenMP 3.0

## Example – parallel pointer chasing on multiple lists using tasks

```
#pragma omp parallel
{
  #pragma omp for private(p)
  for ( int i =0; i <numlists ; i++) {
    p = listheads [ i ] ;
    while (p ) {
      #pragma omp task
      process (p);
      p=next (p ) ;
    }
  }
}
```

76

# OpenMP Tutorial

OpenMP 3.0

## Example: postorder tree traversal

```
void postorder(node *p) {  
    if (p->left)  
        #pragma omp task  
        postorder(p->left);  
    if (p->right)  
        #pragma omp task  
        postorder(p->right);  
    #pragma omp taskwait // wait for descendants  
    process(p->data);  
}
```

Task scheduling point

- Parent task suspended until children tasks complete

77

OpenMP 3.0

## Task switching

- Certain constructs have task scheduling points at defined locations within them
- When a thread encounters a task scheduling point, it is allowed to suspend the current task and execute another (called *task switching*)
- It can then return to the original task and resume

78

# OpenMP Tutorial

OpenMP 3.0

## Task switching example

```
#pragma omp single
{
  for (i=0; i<ONEZILLION; i++)
    #pragma omp task
      process(item[i]);
}
```

- Too many tasks generated in an eye-blink
- Generating task will have to suspend for a while
- With task switching, the executing thread can:
  - ◆ execute an already generated task (draining the “*task pool*”)
  - ◆ dive into the encountered task (could be very cache-friendly)

79

OpenMP 3.0

## Thread switching

```
#pragma omp single
{
  #pragma omp task untied
  for (i=0; i<ONEZILLION; i++)
    #pragma omp task
      process(item[i]);
}
```

- Eventually, too many tasks are generated
- Generating task is suspended and executing thread switches to a long and boring task
- Other threads get rid of all already generated tasks, and start starving...
- With thread switching, the generating task can be resumed by a different thread, and starvation is over
- Too strange to be the default: the programmer is responsible!

80

# OpenMP Tutorial

OpenMP 3.0

## Dealing with taskprivate data

- The Taskprivate directive was removed from OpenMP 3.0
  - ◆ Too expensive to implement
- Restrictions on task scheduling allow threadprivate data to be used
  - ◆ User can avoid thread switching with tied tasks
  - ◆ Task scheduling points are well defined

81

OpenMP 3.0

## Conclusions on tasks

- Enormous amount of work by many people
- Tightly integrated into 3.0 spec
- Flexible model for irregular parallelism
- Provides balanced solution despite often conflicting goals
- Appears that performance can be reasonable

82

# OpenMP Tutorial

OpenMP 3.0

## Nested parallelism

- Better support for nested parallelism
- Per-thread internal control variables
  - ◆ Allows, for example, calling `omp_set_num_threads()` inside a parallel region.
  - ◆ Controls the team sizes for next level of parallelism
- Library routines to determine depth of nesting, IDs of parent/grandparent etc. threads, team sizes of parent/grandparent etc. teams

```
omp_get_active_level()  
omp_get_ancestor(level)  
omp_get_teamsize(level)
```

83

OpenMP 3.0

## Parallel loops

- Guarantee that this works ... i.e. that the same schedule is used in the two loops:

```
!$omp do schedule(static)  
do i=1,n  
  a(i) = ....  
end do  
!$omp end do nowait  
!$omp do schedule(static)  
do i=1,n  
  .... = a(i)  
end do
```

84

# OpenMP Tutorial

OpenMP 3.0

## Loops (cont.)

- Allow collapsing of perfectly nested loops

```
!$omp parallel do collapse(2)
do i=1,n
  do j=1,n
    .....
  end do
end do
```

- Will form a single loop and then parallelize that

85

OpenMP 3.0

## Loops (cont.)

- Made `schedule(runtime)` more useful
  - ◆ can get/set it with library routines

```
omp_set_schedule()
omp_get_schedule()
```
  - ◆ allow implementations to implement their own schedule kinds
- Added a new schedule kind `AUTO` which gives full freedom to the runtime to determine the scheduling of iterations to threads.
- Allowed C++ Random access iterators as loop control variables in parallel loops

86

# OpenMP Tutorial

OpenMP 3.0

## Portable control of threads

- Added environment variable to control the size of child threads' stack

`OMP_STACKSIZE`

- Added environment variable to hint to runtime how to treat idle threads

`OMP_WAIT_POLICY`

`ACTIVE` keep threads alive at barriers/locks

`PASSIVE` try to release processor at barriers/locks

87

OpenMP 3.0

## Control program execution

- Added environment variable and runtime routines to get/set the maximum number of active levels of nested parallelism

`OMP_MAX_ACTIVE_LEVELS`

`omp_set_max_active_levels()`

`omp_get_max_active_levels()`

- Added environment variable to set maximum number of threads in use

`OMP_THREAD_LIMIT`

`omp_get_thread_limit()`

88

# OpenMP Tutorial

OpenMP 3.0

## Odds and ends

- Allow unsigned ints in parallel for loops
- Disallow use of the original variable as master thread's private variable
- Make it clearer where/how private objects are constructed/destroyed
- Relax some restrictions on allocatable arrays and Fortran pointers
- Plug some minor gaps in memory model
- Allow C++ static class members to be threadprivate
- Improve C/C++ grammar
- Minor fixes and clarifications to 2.5

89

## Exercise 7: tasks in OpenMP

- Consider the program linked.c
  - ◆ Traverses a linked list computing a sequence of Fibonacci numbers at each node.
- Parallelize this program using tasks.
- Compare your solution's complexity to an approach without tasks.

90



# OpenMP Tutorial

## Exercise 8: linked lists the hard way

- Consider the program `linked.c`
  - ◆ Traverses a linked list computing a sequence of Fibonacci numbers at each node.
- Parallelize this program using constructs defined in OpenMP 2.5 (loop worksharing constructs ... i.e. don't use OpenMP 3.0 tasks).
- Once you have a correct program, optimize it.

91 

## Conclusion

- OpenMP 3.0 is a major upgrade ... expands the range of algorithms accessible from OpenMP.

92

# OpenMP Tutorial

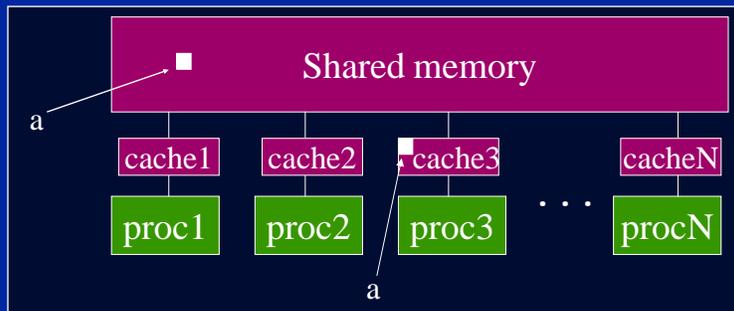
## Outline

- Introduction to OpenMP
- Creating Threads
- Synchronization
- Parallel Loops
- Synchronize single masters and stuff
- Data environment
- OpenMP 3.0 and Tasks
- ➔ • Memory model
- Threadprivate Data

93

## OpenMP memory model

- OpenMP supports a shared memory model.
- All threads share an address space, but it can get complicated:

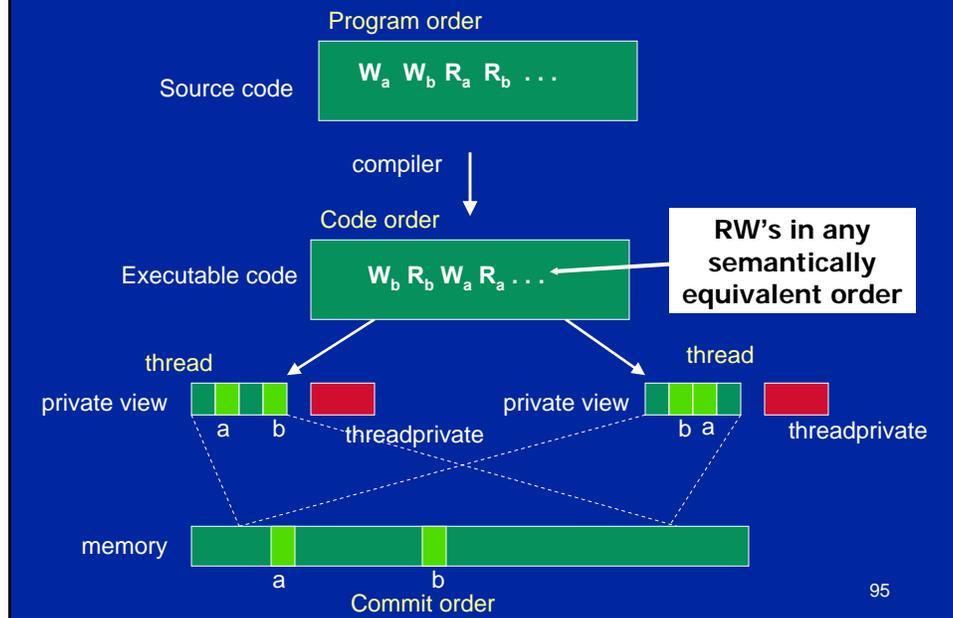


- A memory model is defined in terms of:
  - ◆ **Coherence**: Behavior of the memory system when a single address is accessed by multiple threads.
  - ◆ **Consistency**: Orderings of reads, writes, or synchronizations (RWS) with various addresses and by multiple threads.

94

# OpenMP Tutorial

## OpenMP Memory Model: Basic Terms



## Consistency: Memory Access Re-ordering

- **Re-ordering:**
  - ◆ Compiler re-orders program order to the code order
  - ◆ Machine re-orders code order to the memory commit order
- At a given point in time, the “private view” seen by a thread may be different from the view in shared memory.
- Consistency Models define constraints on the orders of Reads (R), Writes (W) and Synchronizations (S)
  - ◆ ... i.e. how do the values “seen” by a thread change as you change how ops follow ( $\rightarrow$ ) other ops.
  - ◆ Possibilities include:
    - $R \rightarrow R, W \rightarrow W, R \rightarrow W, R \rightarrow S, S \rightarrow S, W \rightarrow S$

96

# OpenMP Tutorial

## Consistency

- **Sequential Consistency:**
  - ◆ In a multi-processor, ops (R, W, S) are sequentially consistent if:
    - They remain in program order for each processor.
    - They are seen to be in the same overall order by each of the other processors.
  - ◆ Program order = code order = commit order
- **Relaxed consistency:**
  - ◆ Remove some of the ordering constraints for memory ops (R, W, S).

97

## OpenMP and Relaxed Consistency

- OpenMP defines consistency as a variant of weak consistency:
  - ◆ S ops must be in sequential order across threads.
  - ◆ Can not reorder S ops with R or W ops on the same thread
    - **Weak consistency guarantees**  
 $S \rightarrow W, S \rightarrow R, R \rightarrow S, W \rightarrow S, S \rightarrow S$
- The Synchronization operation relevant to this discussion is flush.

98

# OpenMP Tutorial

## Flush

- Defines a sequence point at which a thread is guaranteed to see a consistent view of memory with respect to the “flush set”.
- The flush set is:
  - ◆ “all thread visible variables” for a flush construct without an argument list.
  - ◆ a list of variables when the “flush(list)” construct is used.
- The action of Flush is to guarantee that:
  - All R,W ops that overlap the flush set and occur prior to the flush complete before the flush executes
  - All R,W ops that overlap the flush set and occur after the flush don't execute until after the flush.
  - Flushes with overlapping flush sets can not be reordered.

Memory ops: R = Read, W = write, S = synchronization

99

## Synchronization: flush example

- Flush forces data to be updated in memory so other threads see the most recent value

```
double A;  
A = compute();  
flush(A); // flush to memory to make sure other  
          // threads can pick up the right value
```

**Note: OpenMP's flush is analogous to a fence in other shared memory API's.**

100

# OpenMP Tutorial

## Exercise 9: producer consumer

- Parallelize the “prod\_cons.c” program.
- This is a well known pattern called the producer consumer pattern
  - ◆ One thread produces values that another thread consumes.
  - ◆ Often used with a stream of produced values to implement “pipeline parallelism”
- The key is to implement pairwise synchronization between threads.

101

## Exercise 9: prod\_cons.c

```
int main()
{
    double *A, sum, runtime;    int flag = 0;

    A = (double *)malloc(N*sizeof(double));

    runtime = omp_get_wtime();

    fill_rand(N, A);    // Producer: fill an array of data

    sum = Sum_array(N, A); // Consumer: sum the array

    runtime = omp_get_wtime() - runtime;

    printf(" In %lf seconds, The sum is %lf \n",runtime,sum);
}
```

102



# OpenMP Tutorial

## What is the Big Deal with Flush?

- Compilers routinely reorder instructions implementing a program
  - ◆ This helps better exploit the functional units, keep machine busy, hide memory latencies, etc.
- Compiler generally cannot move instructions:
  - ◆ past a barrier
  - ◆ past a flush on all variables
- But it can move them past a flush with a list of variables so long as those variables are not accessed
- Keeping track of consistency when flushes are used can be confusing ... especially if “flush(list)” is used.

Note: the flush operation does not actually synchronize different threads. It just ensures that a thread's values are made consistent with main memory.

103

## Outline

- Introduction to OpenMP
- Creating Threads
- Synchronization
- Parallel Loops
- Synchronize single masters and stuff
- Data environment
- OpenMP 3.0 and Tasks
- Memory model
- ➡ ● Threadprivate Data

104

# OpenMP Tutorial

## Data sharing: Threadprivate

- Makes global data private to a thread
  - ◆ Fortran: **COMMON** blocks
  - ◆ C: File scope and static variables, static class members
- Different from making them **PRIVATE**
  - ◆ with **PRIVATE** global variables are masked.
  - ◆ **THREADPRIVATE** preserves global scope within each thread
- Threadprivate variables can be initialized using **COPYIN** or at time of definition (using language-defined initialization capabilities).

105

## A threadprivate example (C)

Use threadprivate to create a counter for each thread.

```
int counter = 0;
#pragma omp threadprivate(counter)

int increment_counter()
{
    counter++;
    return (counter);
}
```

106

# OpenMP Tutorial

## Data Copying: Copyin

You initialize threadprivate data using a copyin clause.

```
parameter (N=1000)
common/buf/A(N)
!$OMP THREADPRIVATE(/buf/)

C Initialize the A array
call init_data(N,A)

!$OMP PARALLEL COPYIN(A)

... Now each thread sees threadprivate array A initialied
... to the global value set in the subroutine init_data()

!$OMP END PARALLEL

end
```

107

## Data Copying: Copyprivate

Used with a single region to broadcast values of privates from one member of a team to the rest of the team.

```
#include <omp.h>
void input_parameters (int, int); // fetch values of input parameters
void do_work(int, int);

void main()
{
    int Nsize, choice;

    #pragma omp parallel private (Nsize, choice)
    {
        #pragma omp single copyprivate (Nsize, choice)
        input_parameters (Nsize, choice);

        do_work(Nsize, choice);
    }
}
```

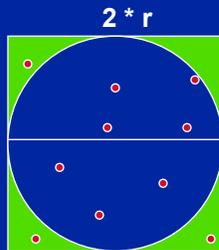
108

# OpenMP Tutorial

## Exercise 10: Monte Carlo Calculations

Using Random numbers to solve tough problems

- Sample a problem domain to estimate areas, compute probabilities, find optimal values, etc.
- Example: Computing  $\pi$  with a digital dart board:



N= 10	$\pi = 2.8$
N=100	$\pi = 3.16$
N= 1000	$\pi = 3.148$

- Throw darts at the circle/square.
- Chance of falling in circle is proportional to ratio of areas:  
 $A_c = r^2 * \pi$   
 $A_s = (2*r) * (2*r) = 4 * r^2$   
 $P = A_c/A_s = \pi / 4$
- Compute  $\pi$  by randomly choosing points, count the fraction that falls in the circle, compute pi.

109

## Exercise 10

- We provide three files for this exercise
  - ◆ pi\_mc.c: the monte carlo method pi program
  - ◆ random.c: a simple random number generator
  - ◆ random.h: include file for random number generator
- Create a parallel version of this program without changing the interfaces to functions in random.c
  - ◆ This is an exercise in modular software ... why should a user of your parallel random number generator have to know any details of the generator or make any changes to how the generator is called?
  - ◆ The random number generator must be threadsafe.
- Extra Credit:
  - ◆ Make your random number generator numerically correct (non-overlapping sequences of pseudo-random numbers).

110

# OpenMP Tutorial

## Conclusion

- We have now covered the full sweep of the OpenMP specification.
  - ◆ We've left off some minor details, but we've covered all the major topics ... remaining content you can pick up on your own.
- Download the spec to learn more ... the spec is filled with examples to support your continuing education.
  - ◆ [www.openmp.org](http://www.openmp.org)
- Get involved:
  - ◆ get your organization to join the OpenMP ARB.
  - ◆ Work with us through Compunity.

111

## Appendices

- ➡ ● Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers

112

# OpenMP Tutorial

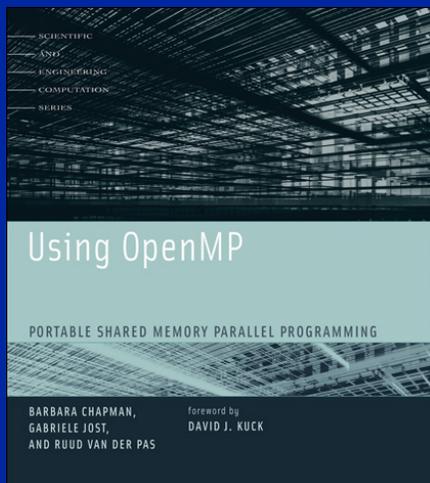
## OpenMP Organizations

- OpenMP architecture review board URL, the “owner” of the OpenMP specification:  
[www.openmp.org](http://www.openmp.org)
- OpenMP User’s Group (cOMPunity) URL:  
[www.compunity.org](http://www.compunity.org)

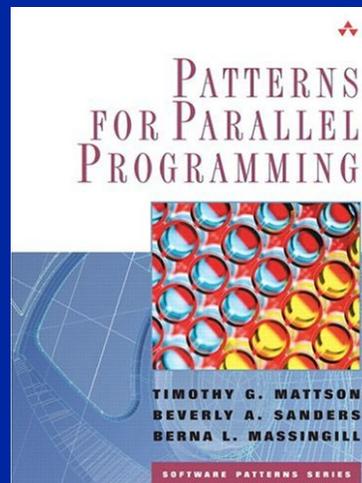
Get involved, join compunity and help define the future of OpenMP

113

## Books about OpenMP



- A new book about OpenMP 2.5 by a team of authors at the forefront of OpenMP’s evolution.



- A book about how to “think parallel” with examples in OpenMP, MPI and java

114

# OpenMP Tutorial

## OpenMP Papers

- Sosa CP, Scalmani C, Gomperts R, Frisch MJ. Ab initio quantum chemistry on a ccNUMA architecture using OpenMP. III. Parallel Computing, vol.26, no.7-8, July 2000, pp.843-56. Publisher: Elsevier, Netherlands.
- Couturier R, Chipot C. Parallel molecular dynamics using OPENMP on a shared memory machine. Computer Physics Communications, vol.124, no.1, Jan. 2000, pp.49-59. Publisher: Elsevier, Netherlands.
- Bentz J., Kendall R., "Parallelization of General Matrix Multiply Routines Using OpenMP", Shared Memory Parallel Programming with OpenMP, Lecture notes in Computer Science, Vol. 3349, P. 1, 2005
- Bova SW, Breshears CP, Cuicchi CE, Demirbilek Z, Gabb HA. Dual-level parallel analysis of harbor wave response using MPI and OpenMP. International Journal of High Performance Computing Applications, vol.14, no.1, Spring 2000, pp.49-64. Publisher: Sage Science Press, USA.
- Ayguade E, Martorell X, Labarta J, Gonzalez M, Navarro N. Exploiting multiple levels of parallelism in OpenMP: a case study. Proceedings of the 1999 International Conference on Parallel Processing. IEEE Comput. Soc. 1999, pp.172-80. Los Alamitos, CA, USA.
- Bova SW, Breshears CP, Cuicchi C, Demirbilek Z, Gabb H. Nesting OpenMP in an MPI application. Proceedings of the ISCA 12th International Conference. Parallel and Distributed Systems. ISCA. 1999, pp.566-71. Cary, NC, USA.

115

## OpenMP Papers (continued)

- Jost G., Labarta J., Gimenez J., What Multilevel Parallel Programs do when you are not watching: a Performance analysis case study comparing MPI/OpenMP, MLP, and Nested OpenMP, Shared Memory Parallel Programming with OpenMP, Lecture notes in Computer Science, Vol. 3349, P. 29, 2005
- Gonzalez M, Serra A, Martorell X, Oliver J, Ayguade E, Labarta J, Navarro N. Applying interposition techniques for performance analysis of OPENMP parallel applications. Proceedings 14th International Parallel and Distributed Processing Symposium. IPDPS 2000. IEEE Comput. Soc. 2000, pp.235-40.
- Chapman B, Mehrotra P, Zima H. Enhancing OpenMP with features for locality control. Proceedings of Eighth ECMWF Workshop on the Use of Parallel Processors in Meteorology. Towards Teracomputing. World Scientific Publishing. 1999, pp.301-13. Singapore.
- Steve W. Bova, Clay P. Breshears, Henry Gabb, Rudolf Eigenmann, Greg Gaertner, Bob Kuhn, Bill Magro, Stefano Salvini. Parallel Programming with Message Passing and Directives; SIAM News, Volume 32, No 9, Nov. 1999.
- Cappello F, Richard O, Etiemble D. Performance of the NAS benchmarks on a cluster of SMP PCs using a parallelization of the MPI programs with OpenMP. Lecture Notes in Computer Science Vol.1662. Springer-Verlag. 1999, pp.339-50.
- Liu Z., Huang L., Chapman B., Weng T., Efficient Implementations of OpenMP for Clusters with Implicit Data Distribution, Shared Memory Parallel Programming with OpenMP, Lecture notes in Computer Science, Vol. 3349, P. 121, 2005

116

# OpenMP Tutorial

## OpenMP Papers (continued)

- B. Chapman, F. Bregier, A. Patil, A. Prabhakar, "Achieving performance under OpenMP on ccNUMA and software distributed shared memory systems," *Concurrency and Computation: Practice and Experience*. 14(8-9): 713-739, 2002.
- J. M. Bull and M. E. Kambites. JOMP: an OpenMP-like interface for Java. Proceedings of the ACM 2000 conference on Java Grande, 2000, Pages 44 - 53.
- L. Adhianto and B. Chapman, "Performance modeling of communication and computation in hybrid MPI and OpenMP applications, *Simulation Modeling Practice and Theory*, vol 15, p. 481-491, 2007.
- Shah S, Haab G, Petersen P, Throop J. Flexible control structures for parallelism in OpenMP; *Concurrency: Practice and Experience*, 2000; 12:1219-1239. Publisher John Wiley & Sons, Ltd.
- Mattson, T.G., How Good is OpenMP? *Scientific Programming*, Vol. 11, Number 2, p.81-93, 2003.
- Duran A., Silvera R., Corbalan J., Labarta J., "Runtime Adjustment of Parallel Nested Loops", *Shared Memory Parallel Programming with OpenMP*, Lecture notes in Computer Science, Vol. 3349, P. 137, 2005

117

## Appendices

- Sources for Additional information
- Solutions to exercises
  - ➡ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers
- Compiler Notes

118

# OpenMP Tutorial

## Exercise 1: Solution

### A multi-threaded “Hello world” program

- Write a multithreaded program where each thread prints “hello world”.

```
#include "omp.h"
void main()
{
#pragma omp parallel
{
    int ID = omp_get_thread_num();
    printf(" hello(%d) ", ID);
    printf(" world(%d) \n", ID);
}
}
```

Annotations:

- OpenMP include file (points to `#include "omp.h"`)
- Parallel region with default number of threads (points to `#pragma omp parallel`)
- Runtime library function to return a thread ID. (points to `omp_get_thread_num()`)
- End of the Parallel region (points to the closing brace of the parallel block)

### Sample Output:

```
hello(1) hello(0) world(1)
world(0)
hello (3) hello(2) world(3)
world(2)
```

119

## Appendices

- Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ➔ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers
- Compiler Notes

120

# OpenMP Tutorial

## The SPMD pattern

- The most common approach for parallel algorithms is the SPMD or Single Program Multiple Data pattern.
- Each thread runs the same program (Single Program), but using the thread ID, they operate on different data (Multiple Data) or take slightly different paths through the code.
- In OpenMP this means:
  - ◆ A parallel region “near the top of the code”.
  - ◆ Pick up thread ID and num\_threads.
  - ◆ Use them to split up loops and select different blocks of data to work on.

121

## Exercise 2: A simple SPMD pi program

```
#include <omp.h>
static long num_steps = 100000;    double step;
#define NUM_THREADS 2
void main ()
{
    int i, nthrds; double pi, sum[NUM_THREADS];
    step = 1.0/(double) num_steps;
    omp_set_num_threads(NUM_THREADS);
    #pragma omp parallel
    {
        int i, id, nthrds;
        double x;
        id = omp_get_thread_num();
        nthrds = omp_get_num_threads();
        if (id == 0) nthrds = nthrds;
        for (i=id, sum[id]=0.0; i< num_steps; i=i+nthrds) {
            x = (i+0.5)*step;
            sum[id] += 4.0/(1.0+x*x);
        }
        for(i=0, pi=0.0; i<nthrds; i++) pi += sum[i] * step;
    }
}
```

Promote scalar to an array dimensioned by number of threads to avoid race condition.

Only one thread should copy the number of threads to the global value to make sure multiple threads writing to the same address don't conflict.

This is a common trick in SPMD programs to create a cyclic distribution of loop iterations

122

# OpenMP Tutorial

## Appendices

- Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ➡ ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers
- Compiler Notes

123

## False sharing

- If independent data elements happen to sit on the same cache line, each update will cause the cache lines to “slosh back and forth” between threads.
  - ◆ This is called “false sharing”.
- If you promote scalars to an array to support creation of an SPMD program, the array elements are contiguous in memory and hence share cache lines.
  - ◆ Result ... poor scalability
- Solution:
  - ◆ When updates to an item are frequent, work with local copies of data instead of an array indexed by the thread ID.
  - ◆ Pad arrays so elements you use are on distinct cache lines.

124

# OpenMP Tutorial

## Exercise 3: SPMD Pi without false sharing

```
#include <omp.h>
static long num_steps = 100000;    double step;
#define NUM_THREADS 2
void main ()
{
    double pi;    step = 1.0/(double) num_steps;
    omp_set_num_threads(NUM_THREADS);
#pragma omp parallel
{
    int i, id, nthrds;    double x, sum; ← Create a scalar local
    id = omp_get_thread_num();    to each thread to
    nthrds = omp_get_num_threads();    accumulate partial
    if (id == 0)    nthrds = nthrds;    sums.
    id = omp_get_thread_num();
    nthrds = omp_get_num_threads();
    for (i=id, sum=0.0; i< num_steps; i=i+nthrds){
        x = (i+0.5)*step;
        sum += 4.0/(1.0+x*x); ← No array, so
    }                                no false
    #pragma omp critical                sharing.
    pi += sum * step; ← Sum goes "out of scope" beyond the
}                                parallel region ... so you must sum it in
}                                here. Must protect summation into pi in
}                                a critical region so updates don't conflict
```

## Appendices

- Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ➔ ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers
- Compiler Notes

126

# OpenMP Tutorial

## Exercise 4: solution

```
#include <omp.h>
static long num_steps = 100000;    double step;
#define NUM_THREADS 2
void main ()
{
    int i;    double x, pi, sum = 0.0;
    step = 1.0/(double) num_steps;
    omp_set_num_threads(NUM_THREADS);
    #pragma omp parallel for private(x) reduction(+:sum)
    for (i=0;i< num_steps; i++){
        x = (i+0.5)*step;
        sum = sum + 4.0/(1.0+x*x);
    }
    pi = step * sum;
}
```

i private  
by default

For good OpenMP  
implementations,  
reduction is more  
scalable than critical.

Note: we created a parallel  
program without changing  
any code and by adding 4  
simple lines!

## Appendices

- Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ➡ ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers
- Compiler Notes

128

# OpenMP Tutorial

## Matrix multiplication

```
#pragma omp parallel for private(tmp, i, j, k)
for (i=0; i<Ndim; i++){
  for (j=0; j<Mdim; j++){
    tmp = 0.0;
    for(k=0;k<Pdim;k++){
      /* C(i,j) = sum(over k) A(i,k) * B(k,j) */
      tmp += *(A+(i*Ndim+k)) * *(B+(k*Pdim+j));
    }
    *(C+(i*Ndim+j)) = tmp;
  }
}
```

- On a dual core laptop
  - 13.2 seconds 153 Mflops one thread
  - 7.5 seconds 270 Mflops two threads

Results on an Intel dual core 1.83 GHz CPU, Intel IA-32 compiler 10.1 build 2

129 

## Appendices

- Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  -  ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers
- Compiler notes

130

# OpenMP Tutorial

## Exercise 6 solution

Compiler will warn you if you have missed some variables

```
#pragma omp parallel for default (none) \  
  shared(x,f,npart,rcoffs,side,sideh) \  
  private(i,j,xi,yi,zi,fxi,fyi,fzi,xx,yy,zz,rd,rrd,rrd2,\  
  rrd3,rrd4,rrd6,rrd7,r148,forcex,forcey,forcez) \  
  reduction(+:epot,vir) \  
  schedule (static,32) \  
  for (i=0; i<npart*3; i+=3) {  
  .....
```

Loop is not well load balanced: best schedule has to be found by experiment.

131

## Exercise 6 solution (cont.)

```
.....  
#pragma omp atomic  
  f[j] -= forcex;  
  ....  
#pragma omp atomic  
  f[j+1] -= forcey;  
  ....  
#pragma omp atomic  
  f[j+2] -= forcez;  
  }  
}  
#pragma omp atomic  
  f[i] += fxi;  
#pragma omp atomic  
  f[i+1] += fyi;  
#pragma omp atomic  
  f[i+2] += fzi;  
  }  
}
```

All updates to f must be atomic

132

# OpenMP Tutorial

## Exercise 6 with orphaning

```
#pragma omp single
{ vir = 0.0;
  epot = 0.0;
}
sideh = 0.5*side;
rcoffs = rcoff*rcoff;
#pragma omp for reduction(+:epot,vir) \
  schedule (static,32)
for (i=0; i<npart*3; i+=3) {
.....
```

Implicit barrier needed to avoid race condition with update of reduction variables at end of the for construct

Each thread initialises private variables here

All variables which used to be shared/private here are now implicitly determined

133

## Exercise 6 with array reduction

```
ftemp[myid][j] -= forcex;
...
ftemp[myid][j+1] -= forcey;
...
ftemp[myid][j+2] -= forcez;
}
}
ftemp[myid][i] += fxi;
ftemp[myid][i+1] += fyi;
ftemp[myid][i+2] += fzi;
}
```

Replace atomics with accumulation into array with extra dimension

134

# OpenMP Tutorial

## Exercise 6 with array reduction

```
....  
#pragma omp for private(i,id)  
  for(i=0;i<(npart*3);i++){  
    for(id=0;id<nthreads;id++){  
      f[i] = f[i] + ftemp[id][i];  
      ftemp[i][id] = 0.0;  
    }  
  }
```

Reduction can be done in parallel

Zero ftemp for next time round

135 

## Appendices

- Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  -  ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers
- Compiler notes

136

# OpenMP Tutorial

## Linked lists with tasks (intel taskq)

- See the file `Linked_intel_taskq.c`

```
#pragma omp parallel
{
  #pragma intel omp taskq
  {
    while (p != NULL) {
      #pragma intel omp task captureprivate(p)
      processwork(p);
      p = p->next;
    }
  }
}
```

	Array, Static, 1	Intel taskq
One Thread	45 seconds	48 seconds
Two Threads	28 seconds	30 seconds

Results on an Intel dual core 1.83 GHz CPU, Intel IA-32 compiler 10.1 build 2

137

## Linked lists with tasks (OpenMP 3)

- See the file `Linked_omp3_tasks.c`

```
#pragma omp parallel
{
  #pragma omp single
  {
    p=head;
    while (p) {
      #pragma omp task firstprivate(p)
      processwork(p);
      p = p->next;
    }
  }
}
```

Creates a task with its own copy of "p" initialized to the value of "p" when the task is defined

138

# OpenMP Tutorial

## Appendices

- Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ➡ ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers
- Compiler notes

139

## Linked lists without tasks

- See the file `Linked_omp25.c`

```
while (p != NULL) {
    p = p->next;
    count++;
}
p = head;
for(i=0; i<count; i++) {
    parr[i] = p;
    p = p->next;
}
#pragma omp parallel
{
    #pragma omp for schedule(static,1)
    for(i=0; i<count; i++)
        processwork(parr[i]);
}
```

Count number of items in the linked list

Copy pointer to each node into an array

Process nodes in parallel with a for loop

	Default schedule	Static,1
One Thread	48 seconds	45 seconds
Two Threads	39 seconds	28 seconds

Results on an Intel dual core 1.83 GHz CPU, Intel IA-32 compiler 10.1 build 2

# OpenMP Tutorial

## Linked lists without tasks: C++ STL

- See the file `Linked_cpp.cpp`

```
std::vector<node*> nodelist;
```

```
for (p = head; p != NULL; p = p->next)
```

```
    nodelist.push_back(p);
```

Copy pointer to each node into an array

```
int j = (int)nodelist.size();
```

Count number of items in the linked list

```
#pragma omp parallel for schedule(static,1)
```

```
    for (int i = 0; i < j; ++i)
```

```
        processwork(nodelist[i]);
```

Process nodes in parallel with a for loop

	C++, default sched.	C++, (static,1)	C, (static,1)
One Thread	37 seconds	49 seconds	45 seconds
Two Threads	47 seconds	32 seconds	28 seconds

Results on an Intel dual core 1.83 GHz CPU, Intel IA-32 compiler 10.1 build 2

141



## Appendices

- Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ➡ ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers
- Compiler Notes

142

# OpenMP Tutorial

## Pair wise synchronizaion in OpenMP

- OpenMP lacks synchronization constructs that work between pairs of threads.
- When this is needed you have to build it yourself.
- Pair wise synchronization
  - ◆ Use a shared flag variable
  - ◆ Reader spins waiting for the new flag value
  - ◆ Use flushes to force updates to and from memory

143

## Exercise 9: producer consumer

```
int main()
{
    double *A, sum, runtime;  int numthreads, flag = 0;
    A = (double *)malloc(N*sizeof(double));

    #pragma omp parallel sections
    {
        #pragma omp section
        {
            fill_rand(N, A);
            #pragma omp flush
            flag = 1;
            #pragma omp flush (flag)
        }
        #pragma omp section
        {
            #pragma omp flush (flag)
            while (flag != 1){
                #pragma omp flush (flag)
            }
            #pragma omp flush
            sum = Sum_array(N, A);
        }
    }
}
```

Use flag to Signal when the  
"produced" value is ready

Flush forces refresh to memory.  
Guarantees that the other  
thread sees the new value of A

Flush needed on both "reader" and  
"writer" sides of the communication

Notice you must put the flush inside the  
while loop to make sure the updated flag  
variable is seen

144



# OpenMP Tutorial

## Appendices

- Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ➡ ◆ Exercise 10: Monte Carlo Pi and random numbers
- Compiler Notes

145

## Computers and random numbers

- We use “dice” to make random numbers:
  - ◆ Given previous values, you cannot predict the next value.
  - ◆ There are no patterns in the series ... and it goes on forever.
- Computers are deterministic machines ... set an initial state, run a sequence of predefined instructions, and you get a deterministic answer
  - ◆ By design, computers are not random and cannot produce random numbers.
- However, with some very clever programming, we can make “pseudo random” numbers that are as random as you need them to be ... but only if you are very careful.
- Why do I care? Random numbers drive statistical methods used in countless applications:
  - ◆ Sample a large space of alternatives to find statistically good answers (Monte Carlo methods).

146

# OpenMP Tutorial

## Monte Carlo Calculations:

### Using Random numbers to solve tough problems

- Sample a problem domain to estimate areas, compute probabilities, find optimal values, etc.
- Example: Computing  $\pi$  with a digital dart board:



N= 10	$\pi = 2.8$
N=100	$\pi = 3.16$
N= 1000	$\pi = 3.148$

- Throw darts at the circle/square.

- Chance of falling in circle is proportional to ratio of areas:

$$A_c = r^2 * \pi$$

$$A_s = (2*r) * (2*r) = 4 * r^2$$

$$P = A_c/A_s = \pi / 4$$

- Compute  $\pi$  by randomly choosing points, count the fraction that falls in the circle, compute pi.

147

## Parallel Programmers love Monte Carlo algorithms

```
#include "omp.h"
static long num_trials = 10000;
int main ()
{
    long i;    long Ncirc = 0;    double pi, x, y;
    double r = 1.0; // radius of circle. Side of square is 2*r
    seed(0,-r, r); // The circle and square are centered at the origin
    #pragma omp parallel for private (x, y) reduction (+:Ncirc)
    for(i=0;i<num_trials; i++)
    {
        x = random();    y = random();
        if ( x*x + y*y) <= r*r) Ncirc++;
    }

    pi = 4.0 * ((double)Ncirc/((double)num_trials);
    printf("\n %d trials, pi is %f \n",num_trials, pi);
}
```

Embarrassingly parallel: the parallelism is so easy its embarrassing.

Add two lines and you have a parallel program.

148

# OpenMP Tutorial

## Linear Congruential Generator (LCG)

- LCG: Easy to write, cheap to compute, portable, OK quality

```
random_next = (MULTIPLIER * random_last + ADDEND)% PMOD;  
random_last = random_next;
```

- If you pick the multiplier and addend correctly, LCG has a period of PMOD.
- Picking good LCG parameters is complicated, so look it up (Numerical Recipes is a good source). I used the following:
  - ◆ MULTIPLIER = 1366
  - ◆ ADDEND = 150889
  - ◆ PMOD = 714025

149

## LCG code

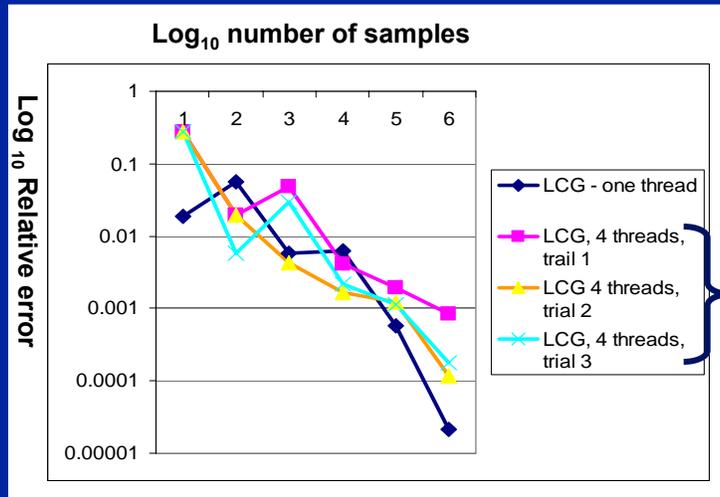
```
static long MULTIPLIER = 1366;  
static long ADDEND    = 150889;  
static long PMOD     = 714025;  
long random_last = 0;  
double random ()  
{  
    long random_next;  
  
    random_next = (MULTIPLIER * random_last + ADDEND)% PMOD;  
    random_last = random_next;  
  
    return ((double)random_next/(double)PMOD);  
}
```

Seed the pseudo random  
sequence by setting  
random\_last

150

# OpenMP Tutorial

## Running the PI\_MC program with LCG generator



Run the same program the same way and get different answers!

That is not acceptable!

Issue: my LCG generator is not threadsafe

Program written using the Intel C/C++ compiler (10.0.659.2005) in Microsoft Visual studio 2005 (8.0.50727.42) and running on a dual-core laptop (Intel T2400 @ 1.83 Ghz with 2 GB RAM) running Microsoft Windows XP.

151

## LCG code: threadsafe version

```
static long MULTIPLIER = 1366;
static long ADDEND    = 150889;
static long PMOD      = 714025;
long random_last = 0;
#pragma omp threadprivate(random_last)
double random ()
{
    long random_next;

    random_next = (MULTIPLIER * random_last + ADDEND)% PMOD;
    random_last = random_next;

    return ((double)random_next/(double)PMOD);
}
```

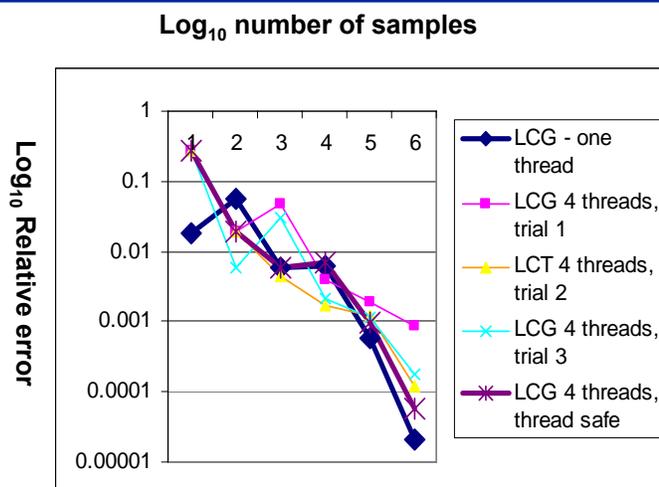
random\_last carries state between random number computations,

To make the generator threadsafe, make random\_last threadprivate so each thread has its own copy.

152

# OpenMP Tutorial

## Thread safe random number generators



Thread safe version gives the same answer each time you run the program.

But for large number of samples, its quality is lower than the one thread result!

Why?

153

## Pseudo Random Sequences

- Random number Generators (RNGs) define a sequence of pseudo-random numbers of length equal to the period of the RNG



- In a typical problem, you grab a subsequence of the RNG range



Seed determines starting point

- Grab arbitrary seeds and you may generate overlapping sequences
  - ◆ E.g. three sequences ... last one wraps at the end of the RNG period.



- Overlapping sequences = over-sampling and bad statistics ... lower quality or even wrong answers!

154

# OpenMP Tutorial

## Parallel random number generators

- Multiple threads cooperate to generate and use random numbers.
- Solutions:
  - ◆ Replicate and Pray
  - ◆ Give each thread a separate, independent generator
  - ◆ Have one thread generate all the numbers.
  - ◆ Leapfrog ... deal out sequence values “round robin” as if dealing a deck of cards.
  - ◆ Block method ... pick your seed so each threads gets a distinct contiguous block.
- Other than “replicate and pray”, these are difficult to implement. Be smart ... buy a math library that does it right.

If done right, can generate the same sequence regardless of the number of threads ...

Nice for debugging, but not really needed scientifically.

Intel’s Math kernel Library supports all of these methods.

155

## MKL Random number generators (RNG)

- MKL includes several families of RNGs in its vector statistics library.
- Specialized to efficiently generate vectors of random numbers

```
#define BLOCK 100
double buff[BLOCK];
VSLStreamStatePtr stream;

vslNewStream(&ran_stream, VSL_BRNG_WH, (int)seed_val);
vdRngUniform (VSL_METHOD_DUNIFORM_STD, stream,
              BLOCK, buff, low, hi);

vslDeleteStream( &stream );
```

Initialize a stream or pseudo random numbers

Select type of RNG and set seed

Fill buff with BLOCK pseudo rand. nums, uniformly distributed with values between lo and hi.

Delete the stream when you are done

156

# OpenMP Tutorial

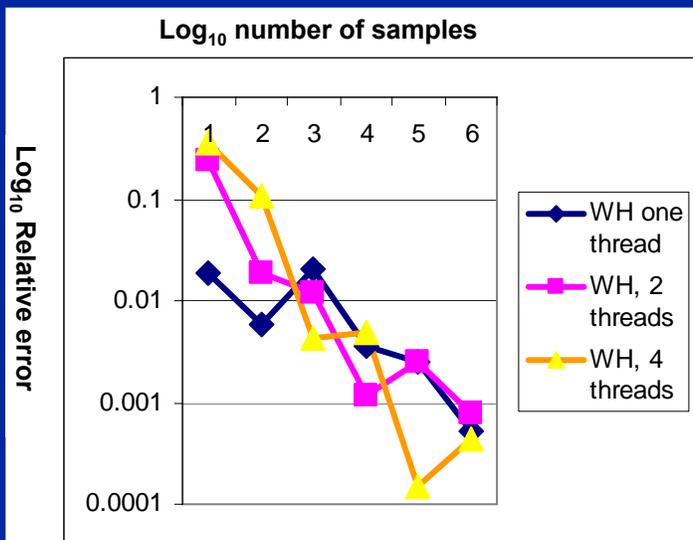
## Wichmann-Hill generators (WH)

- WH is a family of 273 parameter sets each defining a non-overlapping and independent RNG.
- Easy to use, just make each stream threadprivate and initiate RNG stream so each thread gets a unique WG RNG.

```
VSLStreamStatePtr stream;  
#pragma omp threadprivate(stream)  
...  
vsINewStream(&ran_stream, VSL_BRNG_WH+Thrd_ID, (int)seed);
```

157

## Independent Generator for each thread



Notice that once you get beyond the high error, small sample count range, adding threads doesn't decrease quality of random sampling.

158

# OpenMP Tutorial

## Leap Frog method

- Interleave samples in the sequence of pseudo random numbers:
  - ◆ Thread  $i$  starts at the  $i^{\text{th}}$  number in the sequence
  - ◆ Stride through sequence, stride length = number of threads.
- Result ... the same sequence of values regardless of the number of threads.

```
#pragma omp single
{
  nthreads = omp_get_num_threads();
  iseed = PMOD/MULTIPLIER; // just pick a seed
  pseed[0] = iseed;
  mult_n = MULTIPLIER;
  for (i = 1; i < nthreads; ++i)
  {
    iseed = (unsigned long long)((MULTIPLIER * iseed) % PMOD);
    pseed[i] = iseed;
    mult_n = (mult_n * MULTIPLIER) % PMOD;
  }
}
random_last = (unsigned long long) pseed[id];
```

One thread computes offsets and strided multiplier

LCG with Addend = 0 just to keep things simple

Each thread stores offset starting point into its threadprivate "last random" value

## Same sequence with many threads.

- We can use the leapfrog method to generate the same answer for any number of threads

Steps	One thread	2 threads	4 threads
1000	3.156	3.156	3.156
10000	3.1168	3.1168	3.1168
100000	3.13964	3.13964	3.13964
1000000	3.140348	3.140348	3.140348
10000000	3.141658	3.141658	3.141658

Used the MKL library with two generator streams per computation: one for the x values (WH) and one for the y values (WH+1). Also used the leapfrog method to deal out iterations among threads.

160 

# OpenMP Tutorial

## Appendices

- Sources for Additional information
- Solutions to exercises
  - ◆ Exercise 1: hello world
  - ◆ Exercise 2: Simple SPMD Pi program
  - ◆ Exercise 3: SPMD Pi without false sharing
  - ◆ Exercise 4: Loop level Pi
  - ◆ Exercise 5: Matrix multiplication
  - ◆ Exercise 6: Molecular dynamics
  - ◆ Exercise 7: linked lists with tasks
  - ◆ Exercise 8: linked lists without tasks
  - ◆ Exercise 9: Producer-consumer
  - ◆ Exercise 10: Monte Carlo Pi and random numbers

→ • **Compiler Notes**

161

## Compiler notes: Intel on Windows

- Intel compiler:
  - ◆ Launch SW dev environment ... on my laptop I use:
    - start/intel software development tools/intel C++ compiler 11.0/C+ build environment for 32 bit apps
  - ◆ cd to the directory that holds your source code
  - ◆ Build software for program foo.c
    - icl /Qopenmp foo.c
  - ◆ Set number of threads environment variable
    - set OMP\_NUM\_THREADS=4
  - ◆ Run your program
    - foo.exe

To get rid of the pwd on the prompt, type

prompt = %

# OpenMP Tutorial

## Compiler notes: Visual Studio

- Start “new project”
- Select win 32 console project
  - ◆ Set name and path
  - ◆ On the next panel, Click “next” instead of finish so you can select an empty project on the following panel.
  - ◆ Drag and drop your source file into the source folder on the visual studio solution explorer
  - ◆ Activate OpenMP
    - Go to project properties/configuration properties/C.C++/language ... and activate OpenMP
- Set number of threads inside the program
- Build the project
- Run “without debug” from the debug menu.

163

## Compiler notes: Other

- Linux and OS X with gcc:
  - > gcc -fopenmp foo.c
  - > export OMP\_NUM\_THREADS=4
  - > ./a.out
- Linux and OS X with PGI:
  - > pgcc -mp foo.c
  - > export OMP\_NUM\_THREADS=4
  - > ./a.out

for the Bash shell

164